文章编号:1671-4598(2025)09-0310-08 DOI:10.16526/j.cnki.11-4762/tp.2025.09.037 中图分类号:TP391.41 文献标识码:A

基于注意力机制和提示学习的图像去模糊网络

朱金秀[1,2], 徐传蕾[1], 朱京京[1], 苏 新[1]
(1. 河海大学 信息科学与工程学院, 江苏 常州 213022;
2. 海上智能网信技术教育部重点实验室, 江苏 常州 213022)

摘要: 针对现有运动去模糊算法在边缘恢复效果不佳且易产生模糊伪影的问题, 提出了一种基于注意力机制和提示学习的图像去模糊网络; 结合注意力机制设计了特征融合模块, 利用不同层的多尺度信息, 引导网络关注于图像的边缘信息, 以提高图像边缘复原质量; 在解码器中引入轻量级提示模块, 通过捕捉图像的全局结构信息, 增强对模糊区域特征的重建能力, 其中采用两个注意力分支减少了网络参数量和计算量; 实验结果表明, 该网络在3个公开数据集上的定量评价指标均表现优异, 同时参数量和计算量具有一定竞争力, 能够有效恢复图像边缘细节并减少模糊伪影。

关键词: 图像去模糊; 注意力机制; 提示学习; 多尺度网络; 特征融合

Image Deblurring Network Based on Attention Mechanism and Prompt Learning

ZHU Jinxiu[1,2], XU Chuanlei[1], ZHU Jingjing[1], SU Xin[1]
(1. School of Information Science and Engineering, Hohai University, Changzhou 213022, China;
2. Key Laboratory of Maritime Intelligent Cyberspace Technology, Changzhou 213022, China)

Abstract: To address the limitations of existing motion deblurring algorithms, such as poor edge restoration and tendency to produce blurry artifacts, an image deblurring network based on attention mechanism and prompt learning is proposed. A feature fusion module is designed by combining the attention mechanism, which utilizes multi-scale information from different layers to guide the network to focus on edge details, thus improving the quality of edge restoration. A lightweight prompt module is introduced into the decoder to enhance the reconstruction ability of fuzzy region features by capturing the global structure information of the image, of which two attention branches are used to reduce the amount of network parameters and calculations. Experimental results demonstrate that the network performs excellently in quantitative evaluation metrics on three public datasets, while maintaining competitive parameters and computational complexity. The proposed method effectively restores edge details and reduces blurry artifacts.

Keywords: image deblurring; attention mechanism; prompt learning; multi-scale network; feature fusion

0 引言

随着各种图像采集设备的广泛普及, 图像采集变得更加便捷。然而, 由于成像系统的固有限制以及动态场景中环境的不可预测性, 图像模糊成为不可避免的现象, 这不仅显著降低了视觉质量, 还对后续的可视化任务产生不利影响。因此, 图像去模糊是计算机视觉中的一项关键任务。

传统的图像去模糊方法依赖于已知或假设的模糊内核, 并利用有关图像的先验信息进行恢复[1-2]。然而, 这些方法在处理由复杂因素引起的模糊时存在一定的局限性。为了克服非盲去模糊的局限性, 已经开发了端到端的图像去模糊方法。这些方法通过成对的模糊图像和清晰图像来训练网络, 能够直接从模糊图像中恢复清晰的图像, 从而实现较高的去模糊性能。DeepDeblur[3]提

收稿日期:2025-03-20; **修回日期:**2025-05-06。
基金项目:国家自然科学基金项目(62371181);国家重点研发计划(2022YFB4703404);常州市政策引导类计划(国际科技合作/港澳台科技合作 CZ20230029)。
作者简介:朱金秀(1972-),女,博士,教授。
苏 新(1986-),男,博士,教授。
引用格式:朱金秀,徐传蕾,朱京京,等.基于注意力机制和提示学习的图像去模糊网络[J].计算机测量与控制,2025,33(9):310-317,325.

出了一种多尺度架构，利用低尺度模糊图像辅助高尺度图像去模糊，显著提升了图像的恢复效果。然而，这会平滑掉一些应当保留的细节，且参数较多，难以适用于实际场景。SRN[4]采用递归网络减少参数，但该方法容易受到梯度消失问题的限制，同时图像细节恢复效果不足。DeblurGAN-v2[5]利用特征金字塔结构进行特征提取，使得提取的特征更丰富，显著提升了主观视觉效果。然而，该方法未能充分考虑图像模糊区域的非均匀性。DMPHN[6]采用多分块策略学习不同局部特有的模糊运动信息，但这种方法会导致上下文信息的不连续性。MIMO-UNet[7]扩展了多尺度框架，并引入了多输入多输出架构，从不同尺度的模糊图像中提取模糊特征，以恢复出相应的多尺度清晰图像。但在准确捕捉图像边缘信息和全局结构信息方面仍有改进空间。此外，Restormer[8]的 Transformer 架构和 Diffir[9]的扩散模型都在图像去模糊领域取得了显著性能提升。然而，这些架构的引入也带来了较高的模型复杂度和计算开销，限制了其在实时场景中的应用。

针对去模糊网络在图像处理过程中难以有效关注边缘和纹理而导致细节恢复效果欠佳的问题，基于注意力的学习被引入，通过定位目标区域并捕捉其特征[10]，帮助网络更好地聚焦于关键信息，从而提升去模糊效果。Purohit 等[11]使用密集的可变模块和空间注意力机制来处理局部特征，并将这些特征嵌入到完全卷积的设计中，从而实现运动模糊的消除。Chen 等[12]提出了一种针对真实场景的去模糊网络，结合了注意力模块和可变形卷积模块来处理非均匀模糊问题。Ouyang 等[13]提出了一种高效的多尺度注意力机制，克服了降维可能对机器视觉任务带来的负面影响，提升了模型的适应性。由此可见，将注意力机制引入图像去模糊任务，可以使网络聚焦于有助于去模糊的关键特征信息，从而提升去模糊效果。

近期，All-in-One 修复方法逐渐兴起，旨在通过单一模型解决多种图像修复问题，包括去模糊任务。与此同时，提示学习从自然语言处理领域扩展到视觉任务，为 All-in-One 修复提供了新的思路。周等[14]证明，基于条件提示学习的设计在多种任务场景中表现优异。Potlapalli 等[15]通过将其设计的提示块集成到最先进的修复模型中，证明了提示块在图像修复任务中的有效性。Ai 等[16]通过稳定扩散先验，进一步增强了恢复过程，并在去模糊任务中取得了良好的效果。这些研究证明，提示学习使得深度学习模型能够快速适应复杂和动态变化的环境，同时在去模糊等任务中展现出较好的性能。

综上所述，针对现有的运动去模糊算法通常在边缘恢复方面效果不佳，以及缺乏全局结构信息导致模糊伪影的问题，本文提出了一种基于注意力机制和提示学习的图像去模糊网络。该网络结构轻量化，以 MIMO-UNet 为基本架构，结合了注意力与提示学习来提升去模糊效果。首先，设计了一种结合注意力机制的特征融合模块，通过整合多层次尺度特征，引导网络聚焦图像边缘特征；其次，设计了轻量级的提示模块，该模块采用两个注意力分支，在降低参数量的同时，通过多头自注意力分支捕获全局结构信息，并与局部注意力分支融合，强化模糊区域的特征重建能力。这种注意力机制与提示学习结合的设计弥补了单一方法的局限；一方面，提示学习捕捉的全局结构信息有效弥补了注意力机制在长程依赖建模上的不足，显著减少了模糊伪影；另一方面，注意力机制对局部细节的增强作用克服了提示学习在边缘特征提取上的局限性，提升了边缘恢复质量。

实验结果表明，该网络在 3 个公开数据集上的定量评价指标均表现优异，同时参数量和计算量具有一定竞争力，能够有效恢复图像边缘细节并减少模糊伪影。

1　提出的图像去模糊网络

基于注意力机制和提示学习的图像去模糊网络如图 1 所示。网络由 3 个尺度组成，每一个尺度均由 1 个编码块和 1 个解码块组成。首先，网络对输入的模糊图像进行两倍和四倍下采样，生成低分辨率的输入图像。编码块从原始模糊图像中提取特征，同时浅层卷积模块（SCM，shallow convolution module）[7]从低分辨率的模糊图像中提取特征。为了增强边缘复原质量，本文结合特征融合模块与高效多尺度注意力（EMA，efficient multi-scale attention）模块[13]构建了结合注意力机制的特征融合模块（AFFM，attention-integrated feature fusion module）。AFFM 利用多尺度信息引导网络关注图像边缘，从而提升边缘细节的复原精度。此外，本文采用非对称特征整合模块（AFIM，asymmetric feature integration module）[7]融合编码器输出的多尺度特征，以提取更全面的细节信息，避免单一尺度下的特征提取不足。在解码器部分，设计了轻量级提示模块（LPB，lightweight prompt block），以解决网络在图像恢复过程中全局结构信息捕捉不足的问题。LPB 的加入进一步提升了模型在多尺度层次上对全局结构信息的捕捉能力。最终，网络通过卷积层处理，输出去模糊后的清晰图像。

该网络基于 U-Net 结构，主要由 3 个核心模块组成：1）非对称特征整合模块，能够灵活处理不同尺度的特征，并有效地将多尺度特征进行融合；2）结合注意力机制的特征融合模块，引导网络关注于图像的边缘信息，从而提升图像的边缘复原质量；3）轻量级提示模块，结合多头自注意力和局部注意力，增强网络对全

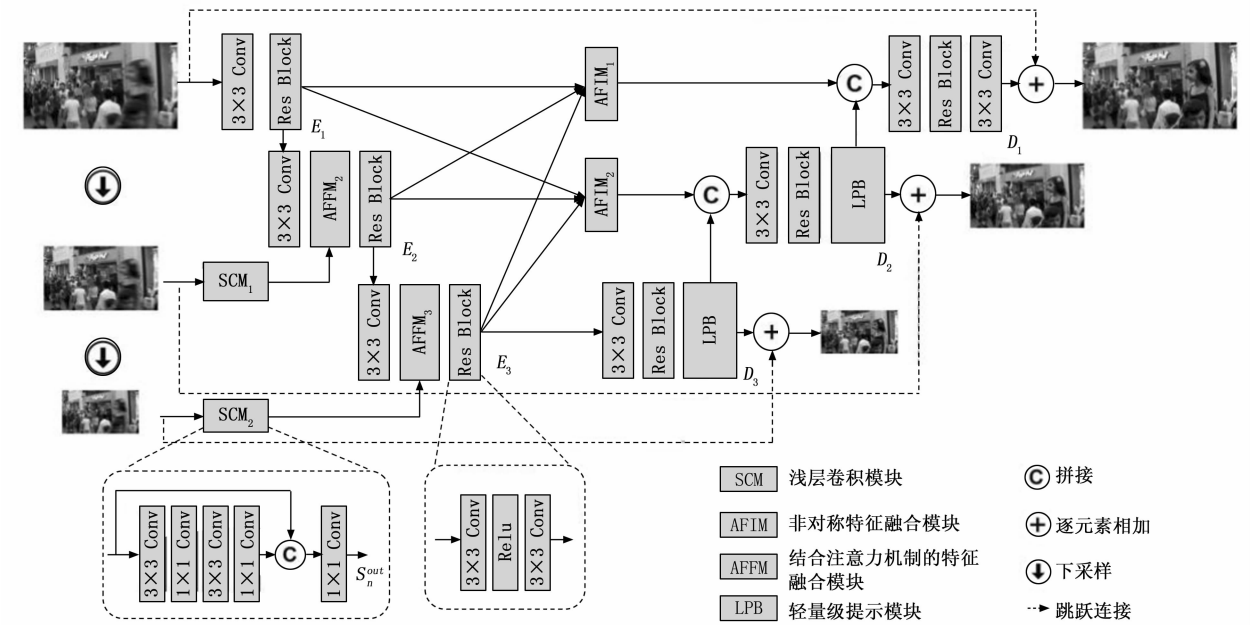

图 1 本文网络结构

局结构和模糊区域特征的恢复能力。

1.1 非对称特征整合模块

网络采用了具有 3 层编码块的特征提取网络，每个 AFIM 都将所有编码块的输出作为输入，通过非对称的上采样和下采样操作动态调整各编码块输出特征的尺度。这种非对称设计充分考虑了不同层次的特征差异，避免了对称融合方法可能造成的信息损失。

在 AFIM 中，经过尺度归一化处理的多层次特征首先进行跨层特征融合，随后通过两个级联的卷积层进行特征增强。最后，增强后的特征将被传递至对应的解码块进行后续处理。第 n 个 AFIM 的输出定义如下：

$$AFIM_n^{\text{out}} = AFIM_n(E_1^{\text{out}}, E_2^{\text{out}}, E_3^{\text{out}}) \tag{1}$$

其中：E_1^{out} 表示第 i 层编码块的输出；$AFIM_n^{\text{out}}$ 表示第 n 个 AFIM 的输出；上采样和下采样操作让来自不同尺度的特征能够融合。因此，解码块能够利用融合后的多尺度特征，有效提高网络的去模糊性能。

1.2 结合注意力机制的特征融合模块

AFFM 如图 2 所示，旨在融合上一层编码块的特征与本层 SCM 输出的特征，为了进一步增强融合特征对边缘信息的关注度，使边缘恢复更加精细，本文在特征融合模块中引入了 EMA 模块[13]。

EMA 模块如图 3 所示。首先，$X \in R^{C\times H\times W}$ 进对输入特征进行特征分组，分为 G 个子特征；来学习不同的边缘特征信息。每组表示为：

$$X = [X_0, X_i, \cdots\cdots X_{G-1}], X_i \in R^{C//G\times H\times W} \tag{2}$$

其中：C 表示输入特征的通道数量，H 和 W 分别表示输入特征的高度和宽度。此外，采用 3 条并行分支提取分组特征图的注意力权重。其中，两个 1×1 分支

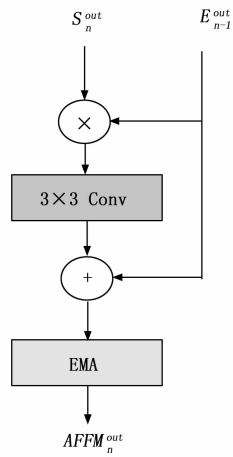

图 2 结合注意力机制的特征融合模块

分别沿垂直和水平方向执行一维全局平均池化，捕捉边缘方向特征，拼接后经 1×1 卷积生成通道注意力并通过 Sigmoid 激活；同时，3×3 分支则通过卷积操作捕获局部边缘特征。通过这种方式，EMA 将精确的空间结构信息保留到通道中。随后进行跨空间信息处理，其中，对 1×1 分支输出应用二维全局平均池化进行编码，池化操作计算公式为：

$$Z_c = \frac{1}{H\times W}\sum_j^H \sum_i^W x_c(i,j) \tag{3}$$

其中：i，j 用来表示像素坐标。通过不同空间维度方向上的跨空间信息聚合，显著提升了对模糊特征的捕获能力，进而为解码器重建图像提供了更丰富的特征信息。

在本文中，将 AFFM 放置在网络的最后两层来进

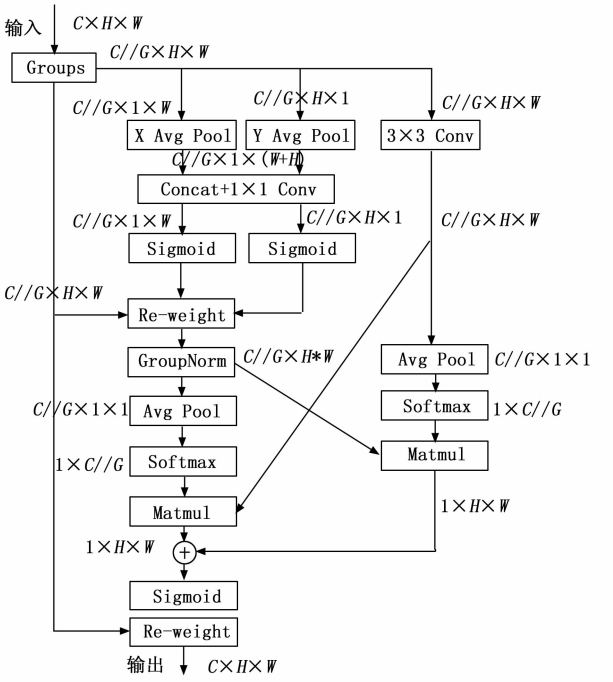

图 3　EMA 模块

行特征处理工作。首先，AFFM 通过元素乘法将上一层编码块提取的特征与本层 SCM 输出的特征进行有效结合。接着，通过卷积层进一步提取多尺度特征，并将其与上一层编码块的特征进行元素加法运算，生成丰富的融合特征。最后，经过 EMA 模块重新校准通道权重，使网络更聚焦于图像边缘信息，从而提升边缘复原精度。AFFM 的流程如下：

$$AFFM_n^{\text{out}} = EMA\left[E_{n-1}^{\text{out}} \oplus Conv_{3\times3}(E_{n-1}^{\text{out}} \otimes S_n^{\text{out}})\right] \tag{4}$$

其中：$S_n{}^{\text{out}}$ 表示在经过第 n 层 SCM 的输出；E_{n-1}^{out} 表示前一个编码块的输出；$AFFM_n^{\text{out}}$ 表示第 n 层的 $AFFM$ 模块；⊕表示元素加法运算；⊗表示元素乘法运算。模型通过引入注意力机制，显著增强了对边缘模糊区域的关注。这种边缘特征有效优化了提示学习的关注区域，使其能够更准确地定位关键重建区域。

1.3　轻量级提示模块

尽管本文模型通过注意力机制实现了对边缘模糊区域的关注，但受限于注意力的局部感受野特性，模型在捕获图像全局结构信息方面仍存在不足，这种局限可能导致重建图像中出现结构失真和伪影现象。为解决这个问题，本文受 promptIR[17] 的启发，提出了一种 LPB，其结构如图 4 所示。该模块通过双分支注意力机制替代传统 Transformer 架构，以实现模型轻量化设计。具体而言，LPB 使用带有多头自注意力（MHSA，multi-head self-attention）的注意力分支捕捉图像的全局结构信息，并与带有局部注意力的注意力分支相融合，从而增强对模糊区域特征的重建能力。

LPB 主要由两个环节组成：条件提示 p 的计算环节和特征交互环节。在 promptIR[17] 中，特征交互环节是基于 Transformer 的。如图 5（a）所示，Transformer[8] 由 MHSA 和门控 Dconv 前馈网络（GDFN，gated-dconv feed-forward network）组成，在图像去模糊任务中表现优秀。其中，MHSA 如图 5（c）所示。首先对输入特征进行层归一化处理，生成特征 Y；随后通过两层卷积得到查询（Q，Query）、键（K，Key）和值（V，Value）；将 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 进行矩阵重塑后得到矩阵 $\hat{\boldsymbol{Q}}$、$\hat{\boldsymbol{K}}$、$\hat{\boldsymbol{V}}$，$\hat{\boldsymbol{Q}}$ 和 $\hat{\boldsymbol{K}}$ 的点积交互生成转置注意图矩阵：

$$\text{Attention}(\hat{\boldsymbol{Q}},\hat{\boldsymbol{K}},\hat{\boldsymbol{V}}) = \hat{\boldsymbol{V}} \cdot \text{Softmax}(\hat{\boldsymbol{K}} \cdot \hat{\boldsymbol{Q}}/\alpha) \tag{5}$$

其中：α 是一个可学习的缩放参数，用于控制和点积结果的幅度。最后，通过特征重塑和卷积操作等完成输出。该方法通过计算通道间的互协方差来生成注意力图，有效增强了全局上下文建模能力，进而提升对图像全局信息的捕捉能力，尤其适用于图像去模糊等任务。但 Transformer 优异的效果伴随着较大的计算量和参数。针对这一问题，本文提出通过两个注意力分支来代替 Transformer，从而实现轻量化。尽管在恢复图像的

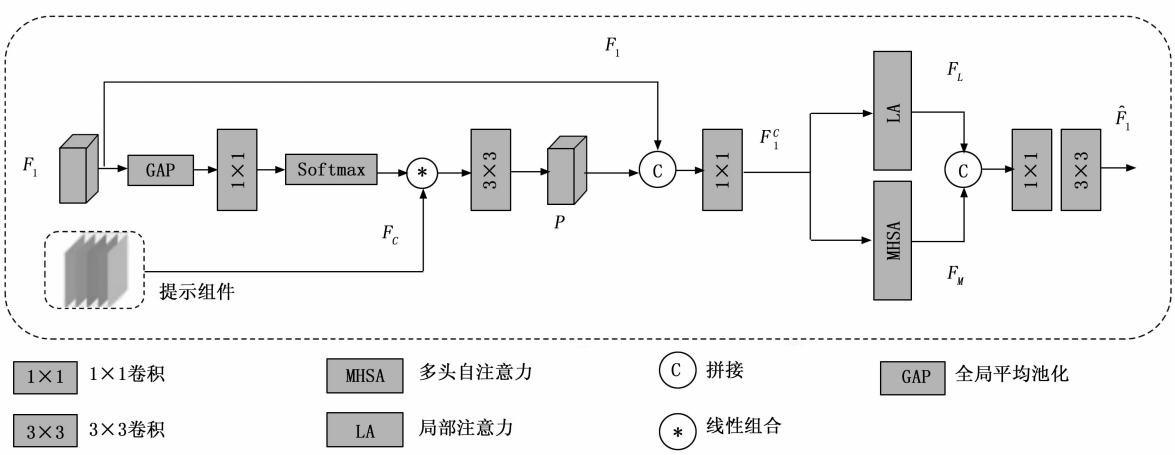

图 4　LPB 的结构

精度上有所损失，但通过减少网络的参数量和计算复杂度，模型整体性能得到了有效提升。对比实验的结果见 2.4 小节。

如图 5（b）所示，本文的两个注意力分支分别是包含 MHSA 的全局注意力分支和包含局部注意力（LA，local attention）的局部注意力分支。首先，为了准确捕捉模糊图像的全局结构信息，MHSA 被保留为全局注意力分支。然后，考虑到大部分运动模糊图像都存在局部严重模糊，局部注意力分支则用于学习局部模糊特征。此外，为了进一步减小网络的参数量和计算量，LA 中使用深度可分离卷积[18]。

在 LPB 中，条件提示 P 的计算环节旨在捕捉丰富的模糊特征。首先，应用全局平均池化在空间维度上进行特征聚合。然后，通过 1×1 卷积层压缩通道数，获得更为紧凑的特征表示。接着，通过 softmax 操作生成提示权重 $\omega\in R^N$，其中 N 的值由提示长度决定，这些权重被用于调整提示组件，以强化对模糊信息的关注。最后，经过 3×3 卷积层生成条件提示 P。综上所述，生成输入条件提示 P 的环节有效增强了模型对模糊信息的处理能力。对于输出特征 $F_1\in R^{H\times W\times C}$，给定 N 个提示组件 $P_c\in R^{N\times H\times W\times C}$，计算条件提示 P 的公式为：

$$P=\mathrm{Conv}_{3\times3}\left(\sum_{c=1}^{N}\omega_i P_c\right),$$
$$\omega_i=\mathrm{Softmax}\{\mathrm{Conv}_{1\times1}[GAP(F_1)]\} \tag{6}$$

其中：GAP 为全局平均池化。特征交互环节旨在增强输入特征 F_1 与条件提示 P 之间的交互，以实现更有效的信息融合。首先，将输入特征 F_1 与条件提示 P 融合，来增强两者之间的交互信息。然后，经过一个卷积层进一步提取特征，得到更加精炼的特征表示 F_1^c，以便更好地挖掘输入特征和提示之间的潜在关系。此步骤如公式（4）所示：

$$F_1^c=\mathrm{Conv}_{1\times1}[Cat(F_1,P)] \tag{7}$$

其中：Cat 表示级联操作；$\mathrm{Conv}_{1\times1}$ 为 1×1 的普通卷积。接下来，生成的特征经过两个分支处理后再融合。在局部注意力分支中，如图 6（e）所示，特征首先经过一个 3×3 的卷积层提取局部模糊信息。然后，经过一个 1×1 的普通卷积并结合残差连接输出特征 F_L，局部注意力分支的整体流程如公式（5）所示：

$$F_L=F_1^c\oplus\mathrm{Conv}_{1\times1}(\mathrm{DSConv}_{3\times3}(F_1^c)) \tag{8}$$

其中：⊕表示元素加法运算，用于特征融合；$\mathrm{Conv}_{1\times1}$ 为 1×1 的普通卷积；$\mathrm{DSConv}_{3\times3}$ 为 3×3 的深度可分离卷积。在全局注意力分支中，特征 F_1^c 经过 MHSA 处理，并结合残差连接输出特征 F_M，输出的特征有丰富的全局结构信息，该分支如公式（6）所示：

$$F_M=F_1^c\oplus\mathrm{MHSA}(F_1^c) \tag{9}$$

其中：F_M 为该分支输出的特征；⊕表示元素加法运算；MHSA 为多头自注意力，如图 6（c）所示。最后，将两个分支的输出结果进行融合，再通过 1×1 卷积和 3×3 卷积的操作，对特征进行处理并映射到输出特征 $\hat{F}_1$。输出特征 $\hat{F}_1$ 为：

$$\hat{F}_1=\mathrm{Conv}_{3\times3}\{\mathrm{Conv}_{1\times1}[Cat(F_L,F_M)]\} \tag{10}$$

其中：Cat 为级联操作。$\mathrm{Conv}_{1\times1}$ 为 1×1 的普通卷积；$\mathrm{Conv}_{3\times3}$ 为 3×3 的普通卷积。

1.4 损失函数

本文的损失函数采用与 MIMO-UNet[7] 相同的策略，使用多尺度内容损失和多尺度频域重建损失（MS-

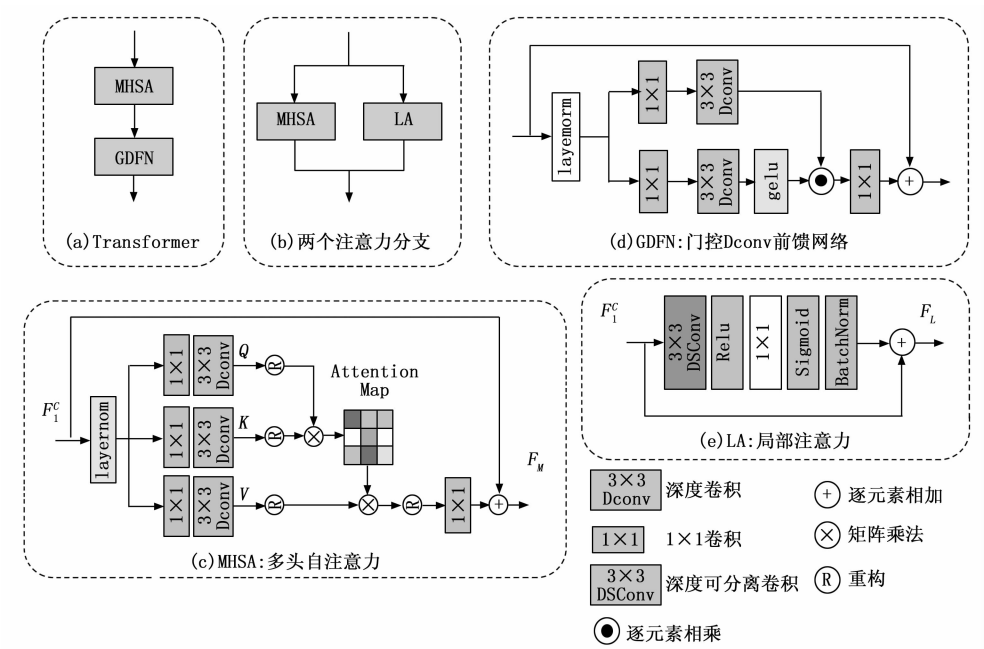

图 5 Transformer 与两个注意力分支

FR，multi-scale frequency reconstruction）。内容损失定义为：

$$L_{\text{cont}} = \sum_{k=1}^{K} \frac{1}{t_k} \| \hat{S}_k - S_k \|_1 \tag{11}$$

其中：K 是级别数；本文将损失除以总元素数 t_k，以进行归一化 $\hat{S}_k$ 为输出图像，S_k 为清晰图像。

除了内容损失外的辅助损失项有助于提高模型性能。由于图像去模糊的目的是恢复丢失的高频部分，因此减少频率空间的差异是非常重要的。因此，损失函数使用了 MSFR 损失函数。MSFR 损失在频域测量多尺度真实图像和去模糊图像之间的 $L1$ 距离，具体公式如下所示：

$$L_{\text{MSFR}} = \sum_{k=1}^{K} \frac{1}{t_k} \| F(\hat{S}_k) - F(S_k) \|_1 \tag{12}$$

其中：F 表示将图像信号传输到频域的快速傅里叶变换。最后，网络训练损失函数定义如下：

$$L_{\text{total}} = L_{\text{cont}} + \lambda L_{\text{MSFR}} \tag{13}$$

其中：λ 是超参数，在实验中设置为 0.1。

2　实验结果和分析

首先，介绍了实验使用的多个数据集和实验细节；接着，对本文方法与竞争方法进行了定量对比试验；最后，进行了定性比较分析与消融实验分析，以进一步验证所提方法的有效性。

2.1　数据集

为了证明本文提出的网络的优越性，本文在 3 个广泛使用的图像去模糊数据集上进行了实验：GoPro 数据集[3]、HIDE 数据集[19]和 RealBlur 数据集[20]。

GoPro 数据集[3]包含 3 214 对清晰和模糊的图像，其中 1 111 对用于评估。该数据集通过高速相机捕捉清晰图像，并根据其生成模糊图像。HIDE 数据集[19]是专注于多重运动模糊的合成数据集，多为行人和街道，包含 2 025 个测试图像对。RealBlur 数据集[20]由真实世界的模糊图像组成，包含两个子集：RealBlur-J 和 RealBlur-R。每个测试集由 980 对图像对组成，没有参考对。

2.2　实验细节

本文使用 GoPro 训练数据集训练模型，并将其应用于 HIDE 数据集和 RealBlur 数据集的测试。对于每次训练迭代，输入图像随机裁剪为大小为 256×256 的采样图像，并使用随机翻转和旋转进行数据增强。为了使去模糊网络更好地收敛，模型训练了 3 500 个 epochs。根据以往实验[7-8]的经验，实验的批量大小设置为 16，使用 Adam 优化器进行优化。为了加快初期收敛，学习率初始值设置为，每 500 个 epochs 降低 0.5 倍。所有实验均在两台 NVIDIA Quadro RTX 5000 GPU 上进行，网络模型使用 PyTorch 深度学习框架实现。

2.3　定量对比实验

本文遵循现有的方法，使用峰值信噪比（PSNR，peak signal-to-noise ratio）和结构相似性（SSIM，structural similarity index measure）作为度量指标，评估恢复图像与真实清晰图像之间的相似性。对于 PSNR 和 SSIM，值越高意味着恢复图像与真实图像的相似性越强。此外，参数量、浮点运算量（FLOPs，floating point operations）和运行时间是衡量一个模型复杂度的 3 个重要指标。

如表 1 所示，在 Gopro 数据集上，本文的方法在 SSIM 上达到了最优值，在 PSNR 上达到了次优值。尽管 MSSNet-small[26]在该数据集中 PSNR 达到最优，但其 FLOPs 是本文方法的近 5 倍，显著增加了计算负担，难以满足实际应用中的能效要求。与目前先进的 MIMO-UNet[7]相比，本文方法在 PSNR 和 SSIM 上分别提高了 0.14 dB 和 0.003，而 FLOPs 仅增加了不到 3%。此外，本文模型在参数量和运行时间方面也具有竞争力。实验结果表明，本文方法在保持较低计算复杂度的同时，实现了较高的复原性能。

表 1　Gopro 数据集的定量测试结果和模型的参数量

模型	PSNR	SSIM	参数量/M	FLOPs/G	时间/s
DeepDeblur[3]	29.23	0.916	11.7	336	4.330
SRN[4]	30.26	0.934	6.8	167	1.870
DeblurGAN-v2[5]	29.55	0.934	N/A	N/A	0.350
PSS-NSC[21]	30.92	0.942	2.84	471	1.600
DMPHN[6]	31.20	0.945	21.70	N/A	0.424
SAPHN[22]	31.85	0.948	23.0	N/A	0.340
DBGAN[23]	31.10	0.942	N/A	N/A	N/A
Attentive deep network[24]	31.23	0.945	26.34	N/A	N/A
MTRNN[25]	31.15	0.945	**2.60**	164	0.070
MSSNet-small[26]	**32.02**	0.953	6.75	317	0.104
MRDNet[27]	31.79	0.951	7.10	72.66	0.677
MIMO-Unet[7]	31.73	0.951	6.81	**66.95**	0.008
本文网络	31.87	**0.954**	7.37	67.73	0.228

注：计算量最佳和次佳结果将突出显示并下划线所有结果均来自现有文献。

为了进一步验证本文方法的去模糊能力与鲁棒性，本文在合成数据集 HIDE 和真实场景数据集 RealBlur 上与其它网络进行对比。从表 2 中可以看出，本文方法在合成数据集 HIDE 和真实场景数据集 RealBlur 上的表现均优于目前先进的 MIMO-UNet[7]。具体而言，在 HIDE 数据集上，本文方法在 PSNR 上提高了 0.52 dB，SSIM 提高了 0.009，这验证了本文方法在复杂运动模糊场景中的去模糊能力；在更具挑战性的真实场景数据集 RealBlur 上的实验进一步证实了本文方法的鲁棒性，特别是在亮度较低的 RealBlur-R 子集上，本文方法在

PSNR 和 SSIM 上都达到了次优值，验证了本文方法在真实场景下的去模糊能力和较强的鲁棒性。

表 2 HIDE 和 RealBlur 数据集的定量测试结果

模型	HIDE		RealBlur-R		RealBlur-J	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DeepDeblur[3]	N/A	N/A	32.51	0.841	27.87	0.827
SRN[4]	28.36	0.915	N/A	N/A	N/A	N/A
DeblurGAN-v2[5]	26.61	0.875	35.26	0.944	**28.70**	**0.866**
DMPHN[6]	29.09	0.924	**35.70**	0.948	28.42	0.860
DBGAN[23]	28.94	0.915	33.78	0.909	24.93	0.745
SAPHN[22]	**29.98**	**0.930**	N/A	N/A	N/A	N/A
MTRNN[25]	29.15	0.918	N/A	N/A	N/A	N/A
MRDNet[27]	29.36	0.921	35.65	**0.951**	28.21	0.862
MIMO-Unet[7]	29.28	0.921	35.47	0.946	27.76	0.837
本文网络	29.80	**0.930**	35.66	0.949	28.30	0.862

注：最佳和次佳结果将突出显示并下划线所有结果均来自现有文献。

2.4 定性对比实验

为了直观展示所提出方法的去模糊效果，本文将其在 GoPro 和 HIDE 数据集上的表现与先进的去模糊方法进行了对比。GoPro 和 HIDE 数据集的视觉比较结果分别如图 6 和图 7 所示。与之前的去模糊方法相比，本文方法能够恢复出更清晰的边缘，并且恢复出更多的图像细节。例如，在图 6 中的 GoPro 数据集的第一行和图 7 中的 HIDE 数据集的第一行中，本文方法能够更清晰地恢复出文本的轮廓信息，且没有产生模糊伪影。相比之下，DeblurGAN-v2[5] 在处理背景严重模糊以及细节丰富的图像时，无法准确提取模糊区域的边界信息，导致去模糊后的图像出现伪影。然而，本文方法在处理相同的模糊图像时，能够获得更加清晰的边缘和更加丰富的细节，证明了其在去模糊任务中的优越性。这主要得益于 AFFM 和 LPB 的加入增强了网络对图像边缘细节和对全局结构信息的提取能力，从而提高了去模糊性能。

图 6 GoPro 数据集上的示例

从左到右：模糊图像、真实清晰图像，以及分别使用 DeblurGan-v2、DBGAN、MIMO-UNet 和本文网络恢复的图像

图 7 HIDE 数据集上的示例

从左到右：模糊图像、真实清晰图像，以及分别使用 DeblurGan-v2、DBGAN、MIMO-UNet 和本文网络恢复的图像

2.5 消融实验

在本文网络中，基于 MIMO-UNet 架构，添加了 AFFM 和 LPB 模块。AFFM 模块帮助网络将注意力集中在图像的边缘信息上，从而更准确地恢复细节和边缘。LPB 模块则通过关注全局结构信息，减少了模糊伪影的产生，有效提升了图像去模糊的效果。

为了验证所提出模块的有效性，本文以 MIMO-UNet[7] 为基线模型，设计了多个实验来分析 AFFM 和 LPB 模块在网络中的作用。本文在 GoPro 数据集和 HIDE 数据集上进行了消融实验。不同组件的消融实验结果如表 3 所示。

表 3 GoPro 和 HIDE 数据集的消融定量结果

模型	GoPro		HIDE	
	PSNR	SSIM	PSNR	SSIM
Baseline	31.73	0.951	29.28	0.921
Baseline+FFM	31.77	0.952	29.48	0.926
Baseline+LPB	31.82	0.953	29.49	0.926
本文算法	31.87	0.954	29.80	0.930

由表 3 的结果可以看出，所加的模块都是有效的。与基线模型相比，单独加入 AFFM 模块和 LPB 模块均能提高网络性能。当所有模块同时使用时，网络在两个数据集上的各项指标均达到了最佳值，这是因为二者的协同工作，既利用注意力机制增强局部边缘特征，又借助提示学习引入全局结构信息，其中 PSNR 在 GoPro 数据集上提升了 0.14 dB，在 HIDE 数据集上提升了 0.52 dB。这些结果表明，本文提出的网络结构组合能够更好地处理图像模糊。同时，HIDE 数据集的良好表现也证明了该网络在未见数据上的鲁棒性和泛化能力。

此外，为了证明本文的 LPB 更为高效，本文还将包含一般提示模块的网络与本文提出的网络进行了对比。表 4 给出了在 GoPro 数据集上的定量指标对比结果。

表 4　GoPro 数据集上 LPB 与 PB 的消融定量结果

模型	PSNR	SSIM	参数量	FLOPs
Baseline	31.73	0.951	6.81	66.95
Baseline＋PB	31.89	0.954	7.50	68.32
Baseline＋LPB	31.87	0.954	7.37	67.73

由表 4 的结果可以看出，尽管本文提出的 LPB 模块在恢复图像的指标上略有损失，但通过减少网络的参数量和 FLOPs，提升了网络的效率，更适合于实际应用。

3　结束语

本文提出了一种基于注意力机制和提示学习的图像去模糊网络，旨在解决运动去模糊算法通常在边缘恢复方面效果不佳，并且由于缺乏图像的全局结构信息，往往会产生模糊伪影的问题。首先，为了进一步增强边缘恢复的精细度，本文结合注意力机制设计了特征融合模块，该模块能够保留通道级的模糊特征。另一方面，LPB 通过双分支注意力机制替代传统 Transformer 架构，实现了模型轻量化设计。其中，多头自注意力分支负责捕捉图像的全局结构信息，而局部注意力分支则专注于提取细节特征。通过双分支特征融合机制，LPB 能够有效增强对模糊区域的特征重建能力，同时保持较低的计算复杂度。

实验结果表明，在定量实验中，本文算法的图像质量评价指标取得了良好的效果。在定性分析中，本文方法能够从模糊图像中恢复出更多细节信息，尤其是在边缘和纹理区域表现出色。此外，消融实验进一步验证了特征融合模块和轻量级提示模块的有效性。然而，本文还需进一步关注图像的边缘信息和全局结构信息，使恢复出的清晰图像信息更加丰富。未来，本网络可以应用在医学影像分析、行车记录仪图像处理与实时视频处理等多个领域。

参考文献：

[1] FERGUS R, SINGH B, HERTZMANN A, et al. Removing camera shake from a single photograph [M]. Acm Siggraph 2006 Papers, 2006: 787 - 794.

[2] YANG L, JI H. A variational EM framework with adaptive edge selection for blind motion deblurring [C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019: 10167 - 10176.

[3] NAH S, HYUN KIM T, MU LEE K. Deep multi-scale convolutional neural network for dynamic scene deblurring [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 3883 - 3891.

[4] TAO X, GAO H, SHEN X, et al. Scale-recurrent network for deep image deblurring [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8174 - 8182.

[5] KUPYN O, MARTYNIUK T, WU J, et al. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better [C] // In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 8878 - 8887.

[6] ZHANG H, DAI Y, LI H, et al. Deep stacked hierarchical multi-patch network for image deblurring [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5978 - 5986.

[7] CHO S J, JI S W, HONG J P, et al. Rethinking coarse-to-fine approach in single image deblurring [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 4641 - 4650.

[8] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: Efficient transformer for high-resolution image restoration [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 5728 - 5739.

[9] XIA B, ZHANG Y, WANG S, et al. Diffir: Efficient diffusion model for image restoration [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 13095 - 13105.

[10] ZHANG K, LUO W, STENGER B, et al. Every moment matters: Detail-aware networks to bring a blurry image alive [C] //Proceedings of the 28th ACM International Conference on Multimedia, 2020: 384 - 392.

[11] PUROHIT K, RAJAGOPALAN A N. Region-adaptive dense network for efficient motion deblurring [C] //Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34 (7): 11882 - 11889.

[12] CHEN L, SUN Q, WANG F. Attention-adaptive and deformable convolutional modules for dynamic scene deblurring [J]. Information Sciences, 2021, 546: 368 - 377.

[13] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning [C] // ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2023: 1 - 5.

[14] ZHOU K, YANG J, LOY C C, et al. Conditional prompt learning for vision-language models [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 16816 - 16825.

[15] LIN J, ZHANG Z, WEI Y, et al. Improving image restoration through removing degradations in textual representations [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 2866 - 2878.

[16] AI Y, HUANG H, ZHOU X, et al. Multimodal prompt perceiver: Empower adaptiveness generalizability and fidelity for all-in-one image restoration [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 25432 - 25444.

（下转第 325 页）